A propos des schémas d'augmentation universels pour la navigation dans les réseaux

Pierre Fraigniaud, Cyril Gavoille, Adrian Kosowski, Emmanuelle Lebhar, Zvi Lotker

À propos des schémas d'augmentation universels pour la navigation dans les réseaux

Pierre Fraigniaud[1], Cyril Gavoille[2], Adrian Kosowski[3], Emmanuelle Lebhar[1] and Zvi Lotker[4]

[1] *CNRS, LIAFA, Université Paris VII. E-mail: {Pierre.Fraigniaud, Emmanuelle.Lebhar}@liafa.jussieu.fr.*
[2] *LABRI, Université Bordeaux II. E-mail: gavoille@labri.fr.*
[3] *Gdansk University of Technoloy, Poland. E-mail: kosowski@sphere.pl.*
[4] *Ben Gurion University, Israël. E-mail: zvilo@cse.bgu.ac.il*

La notion de *graphes augmentés* a été introduite dans le but d'analyser le phénomène des "six degrés de séparation entre individus" observé empiriquement par le psychologue Milgram dans les années 60. De façon formelle, un graphe augmenté est une paire (G,φ), où G est un graphe et φ une collection de distributions de probabilité $\{\varphi_u, u \in V(G)\}$. Un lien supplémentaire est attribué à chaque noeud $u \in V(G)$, pointant vers un noeud v, appelé contact longue-distance de u. La destination v de ce lien est choisie aléatoirement selon $\Pr\{u \to v\} = \varphi_u(v)$. Dans un graphe augmenté, le routage glouton correspond au processus de routage sans mémoire dans lequel chaque noeud intermédiaire choisit parmi ses voisins (dont le voisin supplémentaire) celui qui est le plus proche de la cible selon la distance mesurée dans le graphe de départ G, et lui envoie le message. Grossièrement, les graphes augmentés ont pour but de modéliser la structure des réseaux sociaux tandis que le routage glouton a pour but de modéliser le procédé de recherche utilisé dans l'expérience de Milgram.

Notre objectif est de concevoir des schémas d'augmentation *universels* efficaces, c'est-à-dire des schémas d'augmentation qui attribuent à tout graphe G une collection de distributions de probabilité φ telle que le routage glouton dans (G,φ) soit rapide. Il est connu que le schéma d'augmentation uniforme φ_{unif} est un schéma universel qui garantit que, pour tout graphe G de n noeuds, le routage glouton dans $(G,\varphi_{\text{unif}})$ s'achève en $O(\sqrt{n})$ étapes en espérance. Notre résultat principal est la conception d'un schéma d'augmentation universel φ tel que le routage glouton dans (G,φ) s'achève en $\tilde{O}(n^{1/3})$ étapes en espérance pour tout graphe G de n noeuds. Nous montrons également que sous l'hypothèse d'un modèle plus restreint, la borne $\sqrt{n}$ ne peut être diminuée.

La version complète de cet article est [1].

Keywords: petit monde, routage, graphe augmenté

1 Introduction

Augmented graphs were formally defined in [3] for the purpose of understanding the "small world phenomenon" which consists in the distributed discovery of very short chains between any two nodes. The concept of augmented graphs has recently gained interest, and gave rise to an abundant literature. We refer to Kleinberg's survey [4] on complex networks for more details on this concept.

Formally, an augmented graph is a pair (G,φ) where G is an n-node connected graph, and φ is a collection of probability distributions $\{\varphi_u, u \in V(G)\}$. Every node $u \in V(G)$ is given an extra link, called a *long range link*, pointing to some node v, called the *long range contact* of u, chosen at random according to φ_u as follows : $\Pr\{u \to v\} = \varphi_u(v)$. The links of the underlying graph G are called *local links*.

Greedy routing in (G,φ) is the oblivious routing protocol where the routing decision taken at the current node u for a message of destination t consists in forwarding the message to the neighbor v of u (being local or long range contact) that is the closest to t *according to the distance in* G. This process assumes that every node has a knowledge of the distances in G, but every node is unaware of the long range links added to G, except its own long range link. Hence the nodes have no notion of the distances in the augmented graph.

The *greedy diameter* of (G,φ) is defined as $\mathrm{diam}(G,\varphi) = \max_{s,t\in V(G)} \mathbb{E}(\varphi,s,t)$, where $\mathbb{E}(\varphi,s,t)$ is the expected number of steps for traveling from s to t using greedy routing in (G,φ). Let $f : \mathbb{N} \to \mathbb{R}$ be a function. An n-node graph G is *f-navigable* if there exists a collection of probability distributions φ such that $\mathrm{diam}(G,\varphi) \le f(n)$. Lots of effort have been devoted to characterize the family of graphs that are polylog(n)-navigable (cf. [4]). For instance, the d-dimensional meshes are $O(\log^2 n)$-navigable for any fixed $d \ge 1$ [3]. More generally, it was proved that all graphs of bounded doubling dimension or bounded growth are polylog(n)-navigable. Similarly, all graphs of bounded treewidth, and more generally all graphs excluding a fixed minor are polylog(n)-navigable. All the augmentation schemes proposed in the aforementioned papers are however specifically designed to apply efficiently to each of the considered classes of graphs.

An augmentation scheme is *universal* if it applies to all graphs. The uniform augmentation scheme consists in adding long-range links whose extremities are chosen uniformly at random among all the nodes in the graph. Peleg noticed that any n-node graph is $O(\sqrt{n})$-navigable using this scheme. To see why, consider the ball B of radius $\sqrt{n}$ centered at the target. The expected number of nodes visited until the long range contact of the current node belongs to B is $n/|B|$, and thus at most $\sqrt{n}$. Once in B, the distance to the target is at most $\sqrt{n}$. Up to our knowledge, this $O(\sqrt{n})$ upper bound was the best known bound for arbitrary graphs until this paper. On the other hand, it was recently proved that a function f such that every n-node graph is f-navigable satisfies $f(n) = \Omega(n^{1/\sqrt{\log n}})$ [2]. A crucial problem in the field of network navigability is to close the gap between these upper and lower bounds for the f-navigability of arbitrary graphs.

1.1 Our results.

We first consider augmentation schemes defined a priori from $n \times n$ matrices $A = (a_{i,j})$, where $a_{i,j}$ is the probability that the node labeled i chooses the node labeled j as its long range contact.

- An augmentation of a labeled graph G assigning the links according to the distribution given by A is said *name-independent* since there is no relationship between the labels of A and the ones of G. We prove that, for name-independent schemes, the uniform matrix is optimal in the sense that, for any $n \times n$ matrix A, there is a node labeling of the n-node path P from 1 to n such that greedy routing in (P,A) performs in $\Omega(\sqrt{n})$ expected number of steps.
- To overcome the inefficiency of name-independent schemes, even for graphs as simple as paths, we consider matrix-based augmentation schemes using specific node labelings in $\{1,\ldots,n\}$. We thus consider augmentation schemes φ defined as pairs $(M,\mathcal{L})$ where M is a matrix, and $\mathcal{L}$ is a node-labeling. We describe such a scheme, and analyze its performances in terms of a new parameter, called *pathshape*, achieving tradeoff between pathwidth and pathlength. Precisely, for any n, we describe an $n \times n$ matrix M, and a labeling $\mathcal{L}$ of the nodes of any n-node graph G, such that greedy routing in $(G,(M,\mathcal{L}))$ performs in $O(\min\{\mathrm{ps}(G)\cdot\log^2 n, \sqrt{n}\})$ expected number of steps, where $\mathrm{ps}(G)$ denotes the pathshape of G. In particular, the scheme $(M,\mathcal{L})$ yields polylogarithmic expected number of steps of greedy routing for large classes of graphs such as trees and AT-free graphs, including co-comparability graphs, interval and permutation graphs that were not captured by previous results.

In the second part of the paper, we consider augmentation schemes defined a posteriori, that are fully depending on the structure of the graph. We design a universal augmentation scheme that overcomes the $O(\sqrt{n})$ barrier. Precisely, we design a universal scheme φ such that greedy routing in (G,φ) performs in $\tilde{O}(n^{1/3})$ expected number of steps for any n-node graph G, where the big-$\tilde{O}$ notation ignores the polylogarithmic factors.

2 Matrix-Based Augmentation Schemes

Definition 1. *An* augmentation matrix *of size n is an $n \times n$ matrix $A = (p_{i,j})_{i,j\in[1,n]}$ such that $0 \le p_{i,j} \le 1$ for any $i,j \in [1,n]$, and $\sum_{j=1}^{n} p_{i,j} \le 1$ for any $i \in [1,n]$.*

An augmentation matrix of size n can be used to design augmentation schemes of n-node graphs whose nodes are labeled from 1 to n as follows : node i chooses node j as long range contact with probability $p_{i,j}$.

2.1 Name-Independent Schemes

As we already mentioned in the Introduction, the uniform matrix yields a name-independent augmentation scheme with greedy diameter $O(\sqrt{n})$ for n-node graphs. The following result shows that this is optimal among all matrix-based name-independent augmentation schemes.

Theorem 1. *For any augmentation matrix A of size n, the corresponding name-independent augmentation scheme applied to the n-node path yields greedy diameter $\Omega(\sqrt{n})$.*

Proof sketch. We show that, for any augmentation matrix A of size n, there is a labeling of the n-node path with integers in $\{1,\ldots,n\}$ such that the greedy diameter of the labeled path augmented using A is $\Omega(\sqrt{n})$. Precisely, for any $n\times n$ augmentation matrix $A=(p_{i,j})_{1\leq i,j\leq n}$, we show that there exists a set $I\subseteq\{1,\ldots,n\}$ of cardinality $\sqrt{n}$ such that $\sum_{i,j\in I,i\neq j} p_{i,j} < 1$. Then, assigning the labels of I to $\sqrt{n}$ consecutive nodes on a n-node path, we get that the expected number of long range links on this interval of nodes is strictly less than one. The proof is then completed by showing that this enforces a $\Omega(\sqrt{n})$ greedy diameter for this path. □

The previous result shows that no name-independent scheme can yield greedy diameter better than $\Omega(\sqrt{n})$, even for paths. Yet they remain useful. Indeed, in addition to their simplicity, they can be combined with name-dependent schemes that perform well for specific classes of graphs but poorly in general. Next section proves that the uniform scheme can be combined with a scheme that is efficient for large classes of graphs, in order to preserve the $O(\sqrt{n})$ greedy diameter for general graphs.

2.2 Matrix-Based Augmentation Schemes

In this section, we design a matrix-based augmentation scheme (the matrix is coupled with an appropriate labeling of the nodes) that achieves much better performance than the uniform augmentation scheme for large classes of graphs. Our scheme is based on the new notions of *treeshape* and *pathshape* that establish a tradeoff between the two important notions of treewidth and treelength. These two latter notions have been proved important in many contexts, including algorithm design, routing, and labeling.

Recall that a tree-decomposition of a graph G is a pair (T,X) where T is a tree with node set I of finite size, and $X=\{X_i, i\in I\}$ is a collection of subsets of nodes. T and X must satisfy the three following conditions :

- For any $u\in V(G)$, there exists $i\in I$ for which $u\in X_i$;
- For any $e\in E(G)$, there exists $i\in I$ for which both extremities of e belong to X_i ;
- For any $u\in V(G)$, the set $\{i\in I \mid u\in X_i\}$ induces a subtree of T.

The third constraint can be rephrased as : for any triple $(i,j,k)\in I^3$, if j is on the path between i and k in T, then $X_i\cap X_k\subseteq X_j$. The X_is are called *bags*. When the tree T is restricted to be a path, the resulting decomposition is called path-decomposition. The quality of the tree-decomposition depends on the measure that is applied to the bags X_is. Two measures have been investigated in the past, the width and the length : $width(X_i)=|X_i|-1$, and $\mathrm{length}(X_i)=\max_{x,y\in X_i}\mathrm{dist}_G(x,y)$, where dist_G denotes the distance function in the graph G. (Note that $\mathrm{length}(X_i)$ may be much smaller than the diameter of the subgraph induced by X_i ; In fact X_i may even not be connected). We introduce a new measure, the *shape*, that will be proved very relevant to augmentation schemes.

Definition 2. *The* shape *of a bag X_i of a tree-decomposition (T,X) of a graph G is defined by* $\mathrm{shape}(X_i)=\min\{\mathrm{width}(X_i),\mathrm{length}(X_i)\}$.

The shape of the tree-decomposition is the maximum of the shapes of all its bags. Finally, the treeshape of G (resp., the pathshape of G), denoted by $\mathrm{ts}(G)$ *(resp.,* $\mathrm{ps}(G)$*), is the minimum, taken over all tree-decompositions (resp., path-decompositions) of G, of the shape of the decomposition.*

We show that path-decompositions with small shape can be used to augment efficiently all graphs using a generic matrix and an appropriate labeling that depends on the path-decomposition.

Theorem 2. *For any $n\geq 1$, there exists an $n\times n$ matrix M and a labeling $\mathcal{L}$ of the nodes of any n-node graph G by integers in $\{1,\ldots,n\}$, such that $(G,(M,\mathcal{L}))$ has greedy diameter*

$$O\Big(\min\{\mathrm{ps}(G)\cdot\log^2 n,\ \sqrt{n}\}\Big).$$

Proof sketch. To prove this Theorem, we first construct a labeling of G nodes that describes their relative positions in a path decomposition of G. This labeling is done appropriately so that the bits positions reflects the hierarchy of the decomposition. Then, we consider an augmentation matrix M which is the combination $M = (A+U)/2$ of a matrix A which roughly corresponds to the augmentation matrix of a path along Kleinberg distribution, and of a uniform matrix U. For small pathshape, matrix A combined with the labeling will produce an augmented graph where greedy routing is efficient. But for large pathshape, it turns out that the uniform augmentation remains more efficient, this is the purpose of matrix U. □

An important corollary of Theorem 2 is that the augmentation scheme $(M, \mathcal{L})$ offers a much better behavior than name-independent schemes for large classes of graphs, namely all those having small pathshape. Note that all classes mentioned in the corollary bellow include paths, for which all name-independent augmentation schemes have $\Omega(\sqrt{n})$ greedy diameter. Note also that the mentioned class of AT-free graphs includes co-comporability graphs, interval graphs, and permutation graphs.

Corollary 1. *The universal augmentation scheme of Theorem 2 applied to n-node trees yields greedy diameter $O(\log^3 n)$. Applied to AT-free graphs, it yields greedy diameter $O(\log^2 n)$.*

This corollary is due to the fact that trees have treewidth 1, thus pathwidth at most $O(\log n)$, and hence pathshape at most $O(\log n)$. AT-free graphs have constant pathlength, hence they have pathshape $O(1)$.

As we mentioned in the proof of Theorem 2, nodes may not be assigned different labels by the labeling $\mathcal{L}$. A natural question is whether the label set, and hence the matrix size, could be significantly reduced. The following theorem shows that this is impossible if one wants to preserve polylogarithmic greedy diameter for the classes of graphs mentioned in Corollary 1, or even just for paths. The proof of this Theorem uses similar ideas as Theorem 1 proof.

Theorem 3. *Any matrix-based augmentation-labeling scheme using labels of size $\varepsilon \log n$ for the n-node path, $0 \leq \varepsilon < 1$, yields a greedy diameter $\Omega(n^\beta)$ for any $\beta < \frac{1}{3}(1-\varepsilon)$.*

3 An $\tilde{O}(n^{1/3})$-Step Universal Augmentation Scheme

Neither the uniform scheme nor the augmentation scheme of Theorem 2 enables greedy routing to perform better than $\Omega(n^{1/2})$ expected number of steps for all graphs. The existence of a universal augmentation scheme overcoming the $\Omega(n^{1/2})$ barrier was actually open for some time. In this section, we show that there do exist faster schemes.

Theorem 4. *There exists a universal augmentation scheme φ yielding greedy diameter $\tilde{O}(n^{1/3})$ for n-node graphs.*

Proof sketch. We describe the augmentation scheme φ explicitely. Let G be any (connected) graph. For any node $u \in V(G)$, and any integer $r \geq 0$, let $B(u,r) = \{v \in V(G) \mid \mathrm{dist}_G(u,v) \leq r\}$ be the ball of radius r centered at u. G is augmented as follows. First, every node chooses independently an integer $k \in \{1, \ldots, \lceil \log n \rceil\}$ uniformly at random. Then, the long range contact v of a node u that has chosen integer k is selected uniformly at random in $B_k(u) = B(u, 2^k)$. □

One major open problem in this field is therefore to close the gap between the $\Omega(n^{1/\sqrt{\log n}})$ lower bound, and the $\tilde{O}(n^{1/3})$ upper bound.

Références

[1] P. Fraigniaud, C. Gavoille, A Kosowski, E. Lebhar, and Z. Lotker. Universal Augmentation Schemes for Network Navigability : Overcoming the $\sqrt{n}$-Barrier. In 19th ACM Symp. on Par. Alg. and Archi. (SPAA), 2007.

[2] P. Fraigniaud, E. Lebhar, and Z. Lotker. A Doubling Dimension Threshold $\Theta(\log \log n)$ for Augmented Graph Navigability. In 14th European Symp. on Algorithm (ESA), LNCS 4168, Springer, pages 376-386, 2006.

[3] J. Kleinberg. The Small-World Phenomenon : An Algorithmic Perspective. In 32nd ACM Symp. on Theo. of Comp. (STOC), pages 163-170, 2000.

[4] J. Kleinberg. Complex networks and decentralized search algorithm. In Intl. Congress of Math. (ICM), 2006.